Caliphate Journal of Science & Technology (CaJoST)

ISSN: 2705-313X (PRINT); 2705-3121 (ONLINE)

Research Article

Article Info

Received: 11th December 2019

Revised: 19th January 2020

Accepted: 21st January 2020

[1]*Department of Mathematics, Sokoto State University, Sokoto P.M.B 2134, Sokoto, Nigeria*
[2]*School of Distance Education, Universiti Sains Malaysia, 11800 USM, Pulau Pinang, Malaysia.*
[3]*School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM, Pulau Pinang, Malaysia.*

***Corresponding author's email:**

sas0010@yahoo.com

Cite this: *CaJoST*, 2020, 1, 49-53

3-Satisfiability Reverse Analysis Method with Hopfield Neural Network for Medical Data Set

Samaila A. Mohd, Asyraf Mansor and Saratha Sathasivam

3-Satisfiability Reverse Analysis Method (3SATRA) incorporated with Hopfield neural network is a new approach for the early screening based on the medical dataset. 3SATRA will be proposed to extract the best logic rule that will represent the features of a medical dataset since the conventional data extraction techniques focus only on standalone neural network. The 3-Satisfiability Reverse Analysis Method will be integrated with logic programming as a data mining instrument. The proposed method is applied to Breast Cancer Coimbra and Statlog (Heart) dataset obtained from UCI machine learning repository. To pursue that, the results of the analysis will promote the early screening stage used for medical practitioners. The simulation will be executed using Dev C++ 5.11 as a tool for training, testing and validating the performances of the proposed method. The performance of the method was measured based on Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Sum of Squared Error (SSE), and Computational Time. The performance and accuracy of the results obtained have shown the effectiveness of 3SATRA in medical data mining.

Keywords: 3-Satisfiability; Discrete Hopfield neural network; Medical Dataset; Data mining; Logic programming

1. Introduction

Medical screening is the process of identifying the present or absent of disease and determine the rate at which the disease affects the patients. Hence, accurate detection played a vital role to ensure that the patients were appropriately checked for a certain symptom. The medical specialist is unable to diagnose the disease properly when the number of instances that attributed to the disease increased. Physicians usually apply clinical technologies to assist them in verifying most of the diagnosis issues and give a suggestion as a medical diagnosis is full of uncertainty [1]. The best and efficient way to deals with uncertainty is to apply data mining techniques incorporated with Hopfield neural network and logic mining which are complemented to each other and has the capacity to identify the relationship and cope with the system. To achieved that, data mining model will be formulated to serve as a logic mining tool. In fact, the famous data mining techniques is logic mining.

3-satisfiability problem is specific for general Boolean satisfiability problem which involves constrains on three literals [2]. The 3-Satisfiability (3-SAT) problem will be considered as a mapping problem to assign a logic programming in conjuctive normal form (CNF) to truth values. 3-SAT can be defined as a formula in a conjunctive normal form with a collection of clauses where each comprises or strictly 3 literals per clause [3]. The conjunctive normal form composed of the conjunction of clauses, where clauses are disjunctions of literal [4]. The satisfiability problem can be treated as the combinatorial optimization problem according to logic programming perspective [5]

Logic mining was justifying consistently in mapping as well as representing the attribute of data set into the symbolic logic form [2]. Thus, the major issues with logic mining is the generating logical rule which will be embedded in the Hopfield network or any machined learning models. Therefore 3-satisfiability reversed analysis as a logical rule that has been used to generate the best logic rule and pattern that represent the attribute of the real data set [5]. Furthermore, the robust nature of data mining technique can be hybridized with 3-SAT logic mining. However, the new reform reverses analysis method for extracting valuable information among the attribute of a real data set by using the conjunctive normal form (CNF) logical rule [6].

The conceptual modern-day Hopfield neural network inspired by the human biological nervous system to mock the computations employed by the human brain [7]. The HNN consists of bipolar threshold unit which is full superimposed in dual direction and the connections are symmetric which cause the network to settle to minimum energy [8]. The Hopfield neural network composed of exceptional features such as better stability, recurrent for faster execution and efficient content addressable memory. The discrete HNN computation is performed by collection of activated neurons [9].

The data mining method is used to extract meaningful information of the real data. Based on this paper, the data mining techniques is designed by integrating 3-SAT reverse analysis method (3-SATRA) with Hopfield network were developed to execute the medical data set screening. The method will be simulated using Dev C++ software 5.11. the main idea is to check the performance and efficiency of the proposed method in training, testing and adopting the medical data set as the level of complexities increased.

2. 3-Satisfiability Problem

Hypothetically, three satisfiability (3-SAT) problem can be classified as a classical NP-Hard problem [10]. Satisfiability problem will consider as a task for searching a truth assignment that satisfied logic expression to be true [11]. Generally SAT is a Boolean logic that composed of three literals which allow choices of values for each literals [2]. Hence, three satisfiability (3-SAT) will defined as a mapping problem to assign a Boolean logic in conjunctive normal form (CNF) to truth values. The 3-SAT can be summarized as follows:

I. The 3-Satisfiability in Boolean form comprise set of n variables $q_1, q_2, ..., q_n$ for each clause, for $n=3$.

II. The set of literals which can be positive or negation.

III. The set of s clause in Boolean logic formula, T. $\exists s: T = c_1 \wedge c_2 \wedge, ..., \wedge c_s$ and the clauses are combined with OR operator ($\vee$) and connected with AND operator ($\wedge$).

The precise definition of 3-SAT formula is denoted by:

$$p_{3-SAT} = \bigwedge_{i=1}^{n} c_i \quad (1)$$

where c_i is given by

$$c_i = \bigvee_{i=1}^{n} (x_i, y_i, z_i) \quad (2)$$

Basically, 3-SAT logical rule can be represented as:

$$M = (P \vee \neg Q \vee R) \wedge (\neg S \vee T \vee U) \wedge (V \vee W \neg X) \quad (3)$$

The 3-SAT representation in conjunction normal forma (CNF) is donated by M in equation (3) and the 3-SAT formula can be formulated in various combination as the number of literals assigned at random. The main aimed for the 3-SAT formula is to simplify the output of decision problem.

3. Hopefield Neural Network

The discrete Hopfield neural network is one of the famous neural networks with a wide range of applications. Basically, the Hopfield neural network composed of highly superimposed element (neurons) that form a recurrent HNN. The HNN is fully connected and also feedback to each neurons so that is resemble the biological nervous system [12]. The HNN model is extensively used to performed various optimization problem and as such is formulated by integrating many simple superimposed units (neurons). The HNN consist of good remarkable properties which include parallel execution, faster computation, better stability, and higher memory capacity[13]. The neurons in discrete Hopfield neural network are termed bipolar whereby $s_i = \{-1, 1\}$ which strictly considers values of 1 and -1 [14]. The activation state of neurons is given by:

$$S_i = \begin{cases} 1 \ , \ if = \sum_j R_{ij} S_j > \theta_i \\ -1 \ , \ otherwise \end{cases} \quad (5)$$

where R_{ij} is the synaptic weight from unit i to j. S_j is the state of unit j and θ_i is the threshold of unit i. The connection in Hopfield neural network is symmetric and zero-diagonal $R_{ii} = R_{jj} = 0$ which gives Hopfield symmetrical features in terms of architecture. The updating rule maintains as:

$$S_i(t+1) = \mathrm{sgn}\left[h_i(t)\right] \quad (6)$$

whereby sgn is a sign function of the local field or state of the neuron. According to Equation (6) the updating rule decreases monotonically with the dynamics. The local field crated as:

$$h_i = \sum_j R_{ijk}^{(3)} S_j S_k + \sum_j R_{ij}^{(2)} S_j + R_i^{(1)} \qquad (7)$$

The state of the model develops from initial to final state it will consider local minimum of the energy function. The energy function for the discrete HNN in 3-satisfiability clauses is donated by:

$$E_{Lyapunov} = -\frac{1}{3}\sum_i\sum_j\sum_k R_{ijk}^{(3)} S_i S_j S_k - \frac{1}{2}\sum_i\sum_j R_{ij}^{(2)} S_i S_j - \sum_i R_i^{(1)} S_i$$

(8)

whereby R_{ij} is the synaptic weight, S_i and S_j refer to the neuron state. This energy function is significant because it establishes the degree of convergence and correctness of the network. The energy value obtained from the equation (5) will be verified as global or local minimum energy. Thus, equation (8) is designed only for 3-SAT in HNN.

4. Implementation

The 3-satisfiability reverse analysis (3-SATRA) method is merged with HNN and 3-SAT for doing logic mining. Thus, the execution of the proposed method will be validated during the training and testing of medical data set. The simulations of the designed method will be carried out two medical data set namely: Breast Cancer Coimbra and Statlog (Heart) dataset obtained from UCI machine learning repository. The procedure of is summarized as follows:

Step 1:
Select the data set to be trained and tested by the proposed method. In this paper, we consider 60% for the training and 40% portion for testing.

Step 2:
Convert all attribute of the date set into bipolar and values to the neuron state that will generate the best logic, P_{best} .

Step 3:
Derive a cost function of 3-SAT logic, E_{3-SAT}. by considering $P = \frac{1}{2}(1+S_P)$ and $\neg P = \frac{1}{2}(1-S_P)$. Hence the neuron state can be verified as $S_P = 1$ or $S_P = -1$.

Step 4:
Check the clauses satisfaction of E_{3-SAT}. the model will be trained by exhaustive search (ES) method and best pattern that represent the data set will be store in CAM of HNN.

Step 5:
Compute the synaptic weight of P_{best} using (wan Abdullah 1993) standard weight management techniques in HNN by comparing the energy function and cost function.

Step 6:
Compute the local field, h_i and apply hyperbolic tangent activation function to determine the final state neurons.

Step 7:
Generate the induced logic, $P_{induced}$ to compare the output of the network and the actual target output from the data set which will verified as Success or Failure.

Step 8:
Compute the corresponding performance evaluation metrics namely, RMSE, MAE, SSE, Accuracy and Computational time of model.

5. Results and Discussion

The platform used in simulating the medical data set is Dev C++ 5.11 for training and testing the proposed method in order to validate the models.

Table 1. Experimental Setup

Model	P_{best}
Operating System	Windows 10
Processor	AMD RYZEN 3, 2.5 Ghz
Memory	4GB RAM
Termination of Criterion	0.001
Number of Trial	100
Number of Neurons	1-8

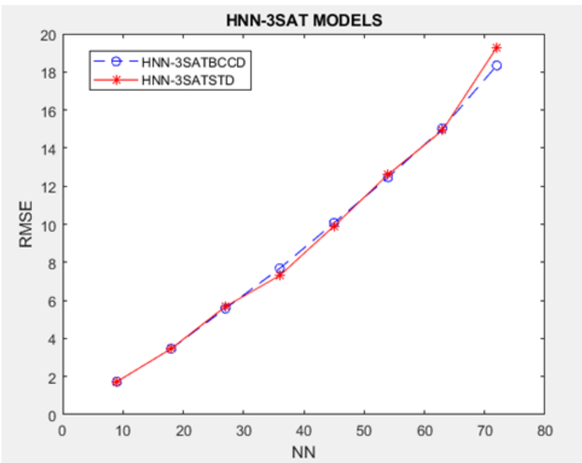

Figure 1. RMSE for HNN-3SATBCCD and HNN-3SATSTD

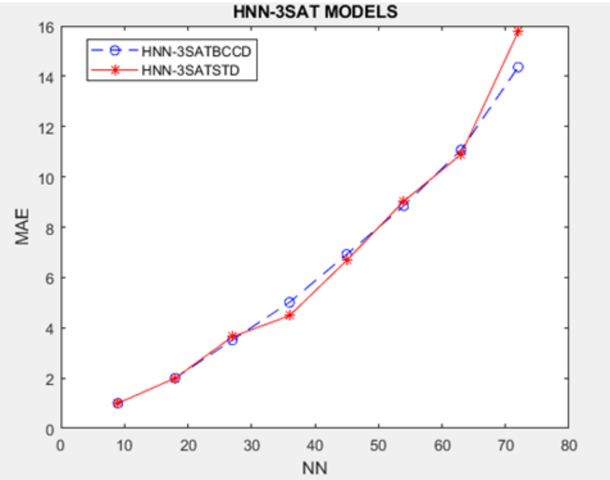

Figure 2. MAE for HNN-3SATBCCD and HNN-3SATSTD

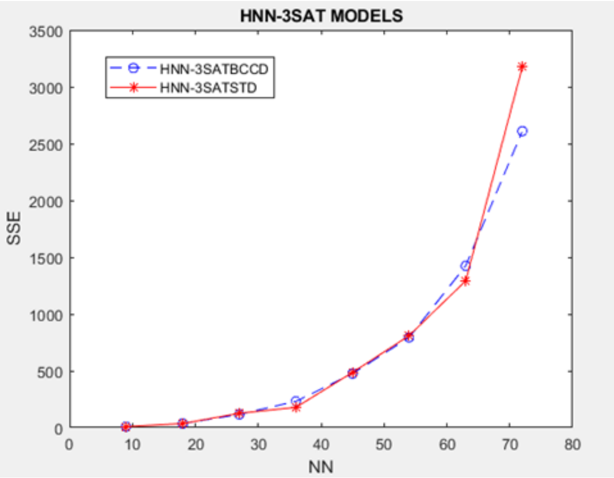

Figure 3. SSE for HNN-3SATBCCD and HNN-3SATSTD

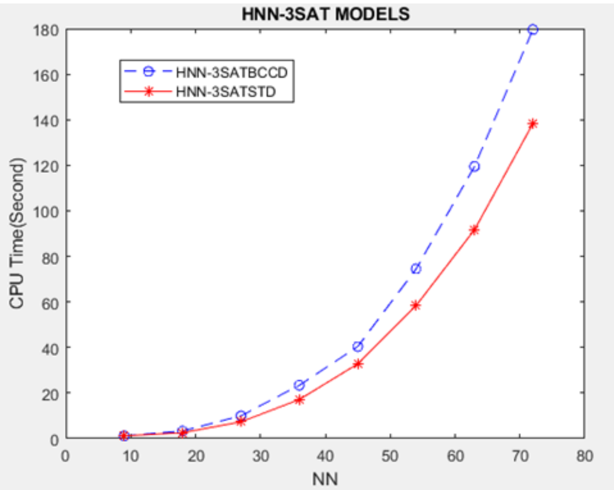

Figure 4. CPU Time for HNN-3SATBCCD and HNN-3SATSTD

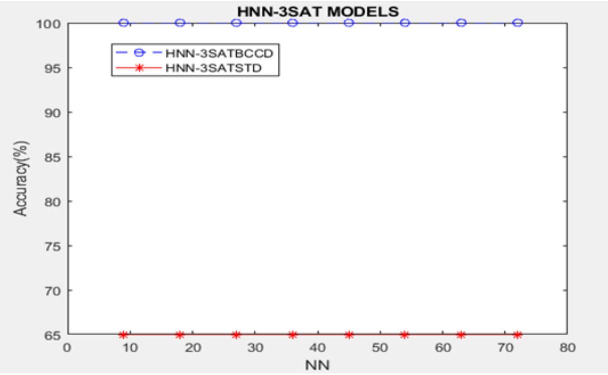

Figure 5. Accuracy for HNN-3SATBCCD and HNN-3SATSTD

The results revealed the performances of HNN-3SATES based on RMSE, MAE, SSE and CPU Time calibration measures in training Breast Cancer Coimbra and Statlog (Heart) data set. In this research, 60% of the data is allocated as the training data set and the remaining 40% as the testing data. The main task is to extract logical rule for Breast Cancer Coimbra and Statlog (Heart) data set. Basically, the proposed network will assist the medical practitioners to identify the presence or absence of the disease during screening. The simulations are carried out until $NC = 8$ for simplicity to check the complexity. the HNN-3SAT model with 3-SATRA has successfully extracted the best logical rule for both data set. Hence, the comparison of the RMSE, MAE, SSE, Accuracy and CPU Time for HNN-3SATBCC and HNN-3SATSD are as follows. Figure 1 depicts the RMSE values obtained by HNN-3SATBCC and HNN-3SATSD for different complexities. Hence, the proposed method was performed consistently better in training the breast cancer Coimbra data set than the other data set. However, the HNN-3SATSD exhibits slightly RMSE than HNN-3SATBCC. The MAE values for HNN-3SATBCC and HNN-3SATSD is shown in Figure 2 there is positive correlation since the MAE values are always positive. The lower the MAE values indicate the solution is quite near to the optimal solution. Therefore, HNN-3SATBCC outperformed HNN-3SATSD during the training phase. In Figure 3 demonstrates the SSE recorded for HNN-3SATBCC and HNN-3SATSD with different array of NC parameters. The error accumulation can be described through the SSE value. Therefore, the magnitude of the SSE will increase as the number of clause increase. Thus, the proposed model performed well in training the Breast Cancer Coimbra and Statlog (Heart) data set. The robustness of our 3-SATRA algorithms can be approximately demonstrated base on the effectiveness of the entire computational process. As the number of neurons increases, the CPU Time also increase for all the hybrid networks. Hence, overall CPU Time when $NC = 8$ the proposed paradigms was generally below 2000 seconds. The ACCURACY recorded for HNN-3SATBCC and HNN-3SATSD were 100% and 65% respectively. The HNN-3SAT with 3-SATRA approach to the Breast cancer Coimbra and Statlog (Heart) data set is proven the performance of the proposed model and it can be implemented by medical practitioners as an early screening tool.

6. Conclusion

The proposed algorithm, HNN-3SATES has demonstrated a good performance based on RMSE, MAE, SSE and CPU time obtained during training and testing process. as such, the proposed hybrid HNN method can be utilized as prediction and classification tool for medical data set. Collectively, the hybrid HNN models with 3-SATRA is proven as a robust paradigm according to the experimental results. Overall, this study has authenticated the capability of our proposed hybrid HNN as a tool in the 3-SAT logic programming and data mining. For future research, the 3-SATRA data mining can be carried out in the other variants of neural network such as convolution neural network (CNN), Wavelet neural network, Deep neural network and a probabilistic neural network. The model can be extended via metaheuristic training method to accelerate classification of complex data set such as streaming and time series data set.

Conflict of interest

The authors declare no conflict of interest.

References

[1] A. M. Sagir and S. Sathasivam, "Intelligence system based classification approach for medical disease diagnosis," *AIP Conf. Proc.*, vol. 1870, 2017.

[2] M. A. Mansor, S. Sathasivam, and M. S. M. Kasihmuddin, "3-satisfiability logic programming approach for cardiovascular diseases diagnosis," in *AIP Conference Proceedings*, 2018.

[3] K. Kutzkov, "New upper bound for the #3-SAT problem," *Inf. Process. Lett.*, 2007.

[4] S. Sathasivam, N. P. Fen, and N. Hamadneh, "Developing agent based modeling for reverse analysis method," *Res. J. Appl. Sci. Eng. Technol.*, 2013.

[5] M. A. Mansor and S. Sathasivam, "Accelerating Activation Function for 3-Satisfiability Logic Programming," *Int. J. Intell. Syst. Appl.*, 2016.

[6] S. Sathasivam and W. A. T. Wan Abdullah, *Comput. (Vienna/New York)*, 2011.

[7] R. Rojas, *Neural Networks: A Systematic Introduction(Springer science & Business Media, Berlin. 2013).*

[8] J. J. Hopfield," *Biophysics (Oxf).*, vol. 81, no. May, pp. 3088–3092, 1984.

[9] M. S. B. M. Kasihmuddin and S. Sathasivam, "Accelerating activation function in higher order logic programming," *AIP Conf. Proc.*, vol. 1750, 2016.

[10] J. L. Johnson, "A neural network approach to the 3-satisfiability problem," *J. Parallel Distrib. Comput.*, 1989.

[11] W. Ahmad Tajuddin Wan Abdullah, *Phys. Lett. A*, 1993.

[12] S. Sathasivam, "Logic Mining in Neural Networks," 2006.

[13] S. Sathasivam, "Application of Neural Networks in Predictive Data Mining," *2011 2nd Int. Conf. Bus. Econ. Res.*, pp. 371–376, 2011.

[14] S. Sathasivam, "Upgrading Logic Programming in Hopfield Network," *Sains Malaysiana*, vol. 39, no. 1, pp. 115–118, 2010.

